

Der Linked-Data-Service der Deutschen Nationalbibliothek: Auslieferung der Metadaten

Version 1.1

Stand: 1. Dezember 2018

Im Auftrag der Deutschen Nationalbibliothek
erstellt von Heike Eichenauer, Jana Hentschke, Jochen Rupp, Dr. Lars G. Svensson

Deutsche Nationalbibliothek (Leipzig, Frankfurt am Main)
2019

<urn:nbn:de:101-2019010906>

Inhaltsverzeichnis

1. Einführung.....	2
2. Direktzugriff	3
2.1. URI-Schema und HTTP-Content-Negotiation.....	3
2.1.1. Entity-URIs.....	3
2.1.2. Description-URIs.....	3
2.1.3. HTTP-Content-Negotiation	3
2.1.4. Non-Serialised-Content-URIs	4
2.1.5. Content-URLs	4
2.2. Legacy-URIs	5
3. Gesamtabzüge.....	5
4. Schnittstellen.....	6
4.1. OAI-Protocol for Metadata Harvesting (OAI-PMH).....	6
4.2. Search/Retrieve via URL (SRU)	6
4.3. Datenshop.....	7

1. Einführung

Die Deutschen Nationalbibliothek¹ (DNB) stellt im Rahmen ihres Linked-Data-Services² ihre Metadaten³ im Format RDF (Resource Description Framework⁴) kostenfrei bereit. Die Gemeinsame Normdatei (GND⁵) und die Titeldaten des Linked-Data-Services sind unter „Creative Commons Zero“-Bedingungen⁶ (CC0 1.0) zur freien Nachnutzung verfügbar. Die RDF-Daten können auf verschiedenen Wegen bezogen werden:

- | **Direktzugriff (s. Abschnitt 2):** Informationen über einzelne Publikationen oder GND-Entitäten, deren Identifikation bekannt ist, können über persistente HTTP-URIs in verschiedenen RDF-Serialisierungen abgerufen werden. HTTP-Content-Negotiation⁷ wird unterstützt. Die ausgelieferten Daten sind aktuell.
- | **Gesamtabzüge (s. Abschnitt 3):** Die Gesamtabzüge sind in verschiedenen RDF-Serialisierungen verfügbar und werden dreimal jährlich aktualisiert⁸.
- | **Schnittstellen (s. Abschnitt 4):**
 - Über die OAI-Schnittstelle⁹ lassen sich DNB-seitig vorgefertigte Kataloge (sets) mittels Abfrage eines definierten Zeitraums nur in der Serialisierung RDF (RDF/XML) beziehen. Die ausgelieferten Daten sind aktuell.
 - Über die SRU-Schnittstelle¹⁰ [nur Serialisierung RDF (RDF/XML)¹¹] und den DNB-Datenshop¹² (verschiedene RDF-Serialisierungen) sind formal oder inhaltlich individuell selektierte Teilmengen der Datensets verfügbar. Die ausgelieferten Daten sind aktuell.

¹ <https://www.dnb.de>

² <https://www.dnb.de/lds>

³ Ausgenommen sind derzeit noch die Daten des Deutschen Buch- und Schriftmuseums und des Deutschen Exilarchivs sowie lokale Normdaten.

⁴ <https://www.dnb.de/rdf>

⁵ <https://www.dnb.de/gnd>

⁶ <http://creativecommons.org/publicdomain/zero/1.0>

⁷ https://de.wikipedia.org/wiki/Content_Negotiation

⁸ <https://www.dnb.de/datendienst>

⁹ <https://www.dnb.de/oai>

¹⁰ <https://www.dnb.de/sru>

¹¹ <https://www.w3.org/TR/rdf-syntax-grammar/>

¹² <https://www.dnb.de/datenshophilfe>

2. Direktzugriff

2.1. URI-Schema und HTTP-Content-Negotiation

Alle Ressourcen werden eindeutig identifiziert über persistente URIs im Namensraum <http://d-nb.info/>.

2.1.1. Entity-URIs

Entitäten wie Publikationen oder GND-Entitäten werden durch sogenannte Entity-URIs identifiziert.

Beispiel:

Der persistente URI <http://d-nb.info/1068193778> identifiziert eindeutig den Roman „Effi Briest“ von Theodor Fontane in der gedruckten 2014er Ausgabe des Reclam-Verlages (DNB-Identifikationsnummer 1068193778).

Beispiel:

Der persistente URI <http://d-nb.info/gnd/118514768> identifiziert eindeutig den Schriftsteller, Theaterregisseur und Dramatiker Bertolt Brecht (GND-Identifikationsnummer 118514768).

Beim Aufruf der Entity-URIs erfolgt immer ein HTTP-Redirect auf die Beschreibung der Ressource (mittels HTTP-Statuscode 303 „See Other“).

2.1.2. Description-URIs

Die Beschreibungen von Publikationen oder GND-Entitäten werden durch URIs identifiziert, die sich von den Entity-URIs durch die Pfadangabe „/about“ unterscheiden. Sie werden als Description-URIs bezeichnet.

Beispiel:

<http://d-nb.info/gnd/118514768/about>
identifiziert die Beschreibung von Bertold Brecht

Beispiel:

<http://d-nb.info/1068193778/about>
identifiziert die Beschreibung des Romans „Effi Briest“ von Theodor Fontane in der gedruckten 2014er Ausgabe des Reclam-Verlages

Beim Aufruf der Description-URIs erfolgt die Auslieferung der Beschreibung der Ressource (mit HTTP-Statuscode 200, „OK“). Wird kein Format und Serialisierung explizit angefordert (s. Abschnitt 2.1.3), wird im Regelfall die HTML-Repräsentation des Datensatzes im Katalog der DNB ausgeliefert.

2.1.3. HTTP-Content-Negotiation

Über das HTTP-Anfrage-Headerfelder „Accept“ kann sowohl an den Entity-URIs als auch den Description-URIs das gewünschte Format und eine Serialisierung angefordert werden. Folgende RDF-Daten werden derzeit unterstützt:

Datenset „DNBTitel“:

- | application/rdf+xml
- | text/turtle
- | application/ld+json
- | application/vnd.dnb.bibframe.rdf+xml (BIBFRAME Prototyp)

Datenset „GND“:

- | application/rdf+xml
- | text/turtle
- | application/ld+json

2.1.4. Non-Serialised-Content-URIs

Identifizierend für die RDF-Beschreibung der Entitäten im Linked-Data-Service stehen URIs mit Pfadangabe „/lds“.

Beispiel:

<http://d-nb.info/gnd/118514768/about/lds>

identifiziert die Beschreibung von Bertold Brecht in RDF.

Beispiel:

<http://d-nb.info/1068193778/about/lds>

identifiziert die Beschreibung des Romans „Effi Briest“ von Theodor Fontane in der gedruckten 2014er Ausgabe des Reclam-Verlages in RDF.

Beim Aufruf dieser Non-Serialised-Content-URIs erfolgt die Auslieferung der RDF-Beschreibung der Ressource (mit HTTP-Statuscode 200, „OK“). Wird keine Serialisierung explizit angefordert (s. Abschnitt 2.1.3), wird standardmäßig die RDF-Turtle-Repräsentation¹³ der Beschreibung ausgeliefert.

2.1.5. Content-URLs

Die RDF-Beschreibungsdokumente sind auch direkt über ihre Standorte abzurufen. Die sogenannten Content-URLs sind nicht persistent, bei Änderungen wird frühzeitig über die Linked-Data-Service-Mailingliste¹⁴ informiert. Derzeit unterscheiden die Content-URLs sich von den Non-Serialised-Content-URIs durch die Dateiendung.

Beispiel:

<http://d-nb.info/gnd/118514768/about/lds.rdf>

liefert die Beschreibung von Bertold Brecht in RDF (RDF/XML).

Beispiel:

<http://d-nb.info/1068193778/about/lds.ttl>

identifiziert die Beschreibung des Romans „Effi Briest“ von Theodor Fontane in der gedruckten 2014er Ausgabe des Reclam-Verlages in RDF (Turtle).

¹³ <https://www.w3.org/TR/turtle/>

¹⁴ <https://lists.dnb.de/mailman/listinfo/lds>

Unterstützt werden folgende Dateiendungen:

- | .rdf für RDF (RDF/XML)
- | .ttl für RDF (Turtle)
- | .jsonld für RDF (JSON-LD)¹⁵

Der RDF-Datendienst „Entity Facts“¹⁶ ist ein Profil des Linked-Data-Services¹⁷. Er ist derzeit ausschließlich über Content-URLs der Form <http://hub.culturegraph.org/entityfacts/<GND-ID>> erreichbar. Zukünftig wird hier voraussichtlich eine Anpassung erfolgen sowie HTTP-Content-Negotiation auf die Dimension „Profil“ eingeführt werden.

2.2. Legacy-URIs

Aus den Anfängen des Dienstes werden noch folgende URIs unterstützt:

- | <http://d-nb.info/.../about/rdf> als Alias für die RDF-Repräsentation RDF (RDF/XML) (abgelöst durch Content-URL <http://d-nb.info/.../about/lds.rdf>)
- | <http://d-nb.info/.../about/bibframe> als Zugang zum BIBFRAME-Prototypen¹⁸ (nur in der Serialisierung RDF (RDF/XML) verfügbar)

3. Gesamtabzüge

Gesamtabzüge der Meta- und Normdaten werden unter <https://data.dnb.de/opendata/> zum kostenfreien Bezug angeboten. Es sind separate Dateien für die Datensets Gemeinsame Normdatei („GND“) und die DNB-Titeldaten („DNBTitel“) verfügbar. Diese werden auch als „Linked-Data-Dumps“ bezeichnet. Wegen ihrer erheblichen Dateigröße werden sie GZIP-komprimiert¹⁹ abgelegt, sind aber auch in diesem Zustand noch mehrere Gigabyte groß. Um die Integrität der Dateien überprüfen zu können, stehen SHA256-Prüfsummen zur Verfügung – diese werden in einer gesonderten Datei am gleichen Ort bereitgestellt.

Die Bereitstellung der GND erfolgt in Paketen pro Entität.

Die Gesamtabzüge werden dreimal jährlich aktualisiert. Die Aktualisierung erfolgt in der Regel im Anschluss an die ILTIS Import/Export Releases²⁰ im Februar, Juni, Oktober. Die Veröffentlichung wird jeweils über die Linked-Data-Service-Mailingliste²¹ bekannt gegeben.

Angeboten werden die Serialisierungen RDF (RDF/XML), RDF (Turtle), RDF (JSON-LD) und RDF (N-Triples) sowie das Dateiformat RDF (HDT)²². Weitere Informationen erhalten Sie auf unserer Webseite²³.

¹⁵ <https://json-ld.org/>

¹⁶ <https://www.dnb.de/entityfacts>

¹⁷ <https://www.dnb.de/lds>

¹⁸ <http://www.dnb.de/bibframe>

¹⁹ <https://de.wikipedia.org/wiki/Gzip>

²⁰ <https://wiki.dnb.de/x/wgcbBQ>

²¹ <http://lists.dnb.de/mailman/listinfo/lds>

²² <http://www.rdfhdt.org/>

²³ <http://www.dnb.de/datendienst>

4. Schnittstellen

Die RDF-Daten sind über die DNB-Schnittstellen OAI²⁴, SRU²⁵ und den Datenshop²⁶ zu beziehen. Die auf diesen Wegen ausgelieferten Metadaten befinden sich auf dem aktuellen zeitlichen Stand. Ein SPARQL-Endpoint ist derzeit nicht verfügbar, für die Zukunft aber geplant. Bis es soweit ist wird auf die Möglichkeit verwiesen, der Linked-Data-Fragments-Schnittstelle mit entsprechenden Clients SPARQL-Anfragen zu stellen.

4.1. OAI-Protocol for Metadata Harvesting (OAI-PMH)

Über die OAI-Schnittstelle²⁷ der DNB lassen sich RDF-Daten in der Serialisierung RDF (RDF/XML) fortlaufend oder innerhalb eines Zeitraums kostenfrei beziehen:

- | vorgefertigte inhaltliche Teilmengen („OAI sets“)
- | Neue und geänderte Datensätze in einem bestimmten Zeitfenster (from – until)

Beispiel:

Die Abfrage

<https://services.dnb.de/oai/repository?verb=ListRecords&from=2018-11-02&until=2018-11-05&metadataPrefix=RDFxml&set=authorities:person>

liefert freigeschalteten Nutzern RDF (RDF/XML)-Beschreibungen der GND-Entitäten vom Typ „Person“, die vom 2. November bis zum 5. November 2018 geändert oder erstellt wurden.

Die Nutzung der OAI-Schnittstelle ist nach Anmeldung und Freischaltung kostenfrei möglich.

4.2. Search/Retrieve via URL (SRU)

Die SRU-Schnittstelle²⁸ der DNB ermöglicht es, durch individuelle Suchanfragen selektierte Datenmengen in RDF (RDF/XML) abzufragen.

Beispiel:

Die Abfrage

<https://services.dnb.de/sru/authorities?version=1.1&operation=searchRetrieve&query=WOE%3DKulturwissenschaften%20and%20MAT=corperations&recordSchema=RDFxml>

liefert freigeschalteten Nutzern RDF (RDF/XML)-Beschreibungen von GND-Entitäten des Typs „Organisation“ in denen das Wort „Kulturwissenschaften“ vorkommt.

Beispiel:

Die Abfrage

<https://services.dnb.de/sru/dnb?version=1.1&operation=searchRetrieve&query=geo%3DXA-DE-SN%20and%20MAT=maps&recordSchema=RDFxml>

liefert freigeschalteten Nutzern RDF (RDF/XML)-Beschreibungen von Karten, die im Bundesland Sachsen erschienen sind.

²⁴ <https://www.dnb.de/oai>

²⁵ <https://www.dnb.de/sru>

²⁶ <https://portal.dnb.de/metadataShop.htm>

²⁷ <https://www.dnb.de/oai>

²⁸ <https://www.dnb.de/sru>

Die Nutzung der SRU-Schnittstelle ist nach Anmeldung und Freischaltung kostenfrei möglich.

4.3. Datenshop

Im Datenshop²⁹ der DNB können Rechercheergebnisse aus dem DNB-Katalog³⁰ in verschiedenen RDF-Serialisierungen bezogen werden. Dies bietet sich für nicht-regelmäßige, manuelle Selektion von Treffermengen überschaubarer Größe an.

Die Nutzung des Datenshops ist nach Anmeldung kostenfrei möglich³¹.

²⁹ <https://portal.dnb.de/metadataShop.htm>

³⁰ <https://portal.dnb.de/>

³¹ <https://portal.dnb.de/myAccount/register.htm>